Technical Report 71-22

HumRRO

AD 739521

Comparison and Evaluation of Printed Programs for Aircraft Recognition

Elmo E. Miller and Arthur C. Vicory

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

October 1971

Prepared for
Office of the Chief of Research and Development
Department of the Army
Washington, D.C. 20310

36

R

BIBLIOGRAPHIC DATA SHEET

1. Report No. HumRRO-TR-71-22

2.

3. Recipient's Accession No.

4. Title and Subtitle

COMPARISON AND EVALUATION OF PRINTED PROGRAMS FOR AIRCRAFT RECOGNITION

5. Report Date October 1971

6.

7. Author(s) Elmo E. Miller and Arthur C. Vicory

8. Performing Organization Rept. No. TR 71-22

9. Performing Organization Name and Address

Human Resources Research Organization (HumRRO)
300 North Washington Street
Alexandria, Virginia 22314

10. Project/Task/Work Unit No. 2Q062107A712

11. Contract/Grant No. DAHC-19-70-C-0012

12. Sponsoring Organization Name and Address

Office, Chief of Research and Development
Department of the Army
Washington, D.C. 22314

13. Type of Report & Period Covered

14.

15. Supplementary Notes

Work Unit STAR, Aircraft Recognition Training, HumRRO Division No. 5, Fort Bliss, Texas

16. Abstracts

Several printed prototype programs for training visual aircraft recognition were developed and compared experimentally. One program produced an average score of 9[illegible]% on a printed recognition test (the next closest group test had more than twice as many errors). The program also tended to take the least time to administer (about 15 minutes per aircraft). The training was in three phases: (1) Study of Multi-Image Cards (each showing several views of one aircraft, listing distinctive features); (2) Study of Paired Comparison cards (each showing two or three aircraft that are likely to be confused; (3) Study of Flash Cards (each showing one view of one aircraft-10 different cards for each aircraft). After each phase, tests with printed imagery were administered. The program should be feasible and effective for routine training.

17. Key Words and Document Analysis. 17a. Descriptors

*Programmed instruction

*Military training

17b. Identifiers/Open-Ended Terms

Aircraft identification training

Aircraft recognition

Forward area air defense identification

17c. COSATI Field/Group 05/09 Behavioral and social sciences/Personnel selection, training, and evaluation

18. Availability Statement

Approved for public release; distribution unlimited.

19. Security Class (This Report) UNCLASSIFIED

20. Security Class (This Page) UNCLASSIFIED

21. No. of Pages 54

22. Price

FORM NTIS-35 (10-70)

USCOMM-DC [illegible]

The Human Resources Research Organization (HumRRO) is a nonprofit corporation established in 1969 to conduct research in the field of training and education. It is a continuation of The George Washington University Human Resources Research Office. HumRRO's general purpose is to improve human performance, particularly in organizational settings, through behavioral and social science research, development, and consultation. HumRRO's mission in work performed under contract with the Department of the Army is to conduct research in the fields of training, motivation and leadership.

The contents of this paper are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Published
October 1971
by
HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street
Alexandria, Virginia 22314

FOREWORD

The interest in visual aircraft recognition has been revived in recent years because of the development of a variety of forward area air defense weapons. This report describes an experiment comparing various printed training programs for aircraft recognition training, and discusses the effectiveness to be expected if the apparent best program were to be applied in routine training.

This research and development effort was conducted by the Human Resources Research Organization under Sub-Unit III of Work Unit STAR. Preceding work under STAR, Sub-Unit I served as a basis for the development of printed materials and training procedures which were compared in the experiment.

The various training techniques which were compared were suggested by various staff members of HumRRO Division 5. The printed materials were developed by Mr. Harold E. Christensen and SP4 James C. McBurney, and preliminary experimentation was conducted by SP4 McBurney under the direction of Dr. Elmo E. Miller. The experimental work reported here was conducted by Dr. Miller, Dr. Arthur C. Vicory, and PFC W. Mark Hall.

STAR research, begun in 1965, is being conducted at HumRRO Div. 5, Fort Bliss, Texas. Dr. Robert D. Baldwin was Director of Research during the period in which the research described in this report was performed. Dr. Albert L. Kubala is the present Director of Research.

Military support has been provided by the U.S. Army Air Defense Human Research Unit and by the U.S. Army Air Defense Center. The Military Chief of the Human Research Unit at the time the study was conducted was LTC Frank R. Husted.

HumRRO research for the Department of the Army is conducted under Contract DAHC 19-70-C-0012. Training, Motivation, and Leadership Research is conducted under Army Project 2Q062107A712.

Meredith P. Crawford
President
Human Resources Research Organization

SUMMARY AND CONCLUSIONS

MILITARY PROBLEM

Visual recognition is used by all crews of forward area air defense weapons for aircraft identification. Aircraft recognition training in the U.S. military services has traditionally consisted of group instruction using projected slide images, and currently a much-improved slide kit, the Ground Observer Aircraft Recognition (GOAR) kit, is under development. However, such an approach needs to be supplemented in many Army units because they also need training materials that can be used for self-study or with very small groups (or individual trainees) on a highly flexible training schedule. These needs could be met with training materials that use printed images. It would be desirable for the printed materials to use the same basic photography as the GOAR kit so that they could be produced with minimal cost as soon as the GOAR kit is developed.

RESEARCH OBJECTIVE

The primary objective of the research was to develop and evaluate an effective and efficient prototype of a printed, self-instructional aircraft recognition training program, to be used as a supplement to the GOAR kit.

RESEARCH APPROACH

Several kinds of prototype printed programs were developed and compared experimentally. The comparative evaluation was designed not only to determine the best program, but also to assess the performance level produced by the best program and the amount of training time required.

RESULTS

One program appeared to be better than any of the others, producing a high level of performance in a modest amount of time. This program produced an average score of approximately 95% on a printed recognition test (the next closest group made more than twice as many errors); in addition, on the GOAR slide test administered after the training, the same group made the highest score (about 87%). This program also tended to take the least time to administer (about 15 minutes per aircraft for the average student).

The apparent best program involved three phases: (a) Study of Multi-Image Cards (each card shows several views of one aircraft and lists its most distinctive features), (b) Study of Paired Comparison cards (each card shows two or three aircraft which are apt to be confused from that viewpoint), and (c) Study of Flash Cards (each card showed one view of one aircraft, and there were ten different cards for each aircraft). After each phase, tests with printed imagery were administered to focus attention on the instructional goal and to measure each man's progress.

The Department of Doctrine Development, Literature and Plans of the U.S. Army Air Defense School has prepared a limited number of printed aircraft recognition materials based upon the prototype program which was administered to the highest scoring group.

CONCLUSION

If the training method used for the best scoring experimental group were to be applied in routine training, a high level of performance would be expected with a modest amount of training time. To achieve these results, however, care must be taken not only to use the same kinds of training materials but *also* the same kind of instructions to students and system of testing. These all form integral parts of the training method.

CONTENTS

		Page
Introduction		3
	Military Problem	3
	Research Problem	3
	General Characteristics of Effective Programs	3
	Previous Applications of the Principles	5
Method		6
	Materials	6
	Experimental Criteria	12
	Experimental Groups	13
	Students	14
	Procedure	14
Results		15
Discussion		22
Literature Cited		27
Appendix		
A	Standard Instructions	29
Tables		
1	Training Programs Administered to the Various Groups	13
2	Mean Test Scores	15
3	Distribution of Scores After Training	17
4	Significance Tests of Differences Between Group I and Other Groups	18
5	Time Required for the Component Procedures	19
Figures		
1	Sample Multi-Image Card	7
2	Sample Paired-Comparison Card	8
3	Sample Flash Card	10
4	Sargeant Materials	11
5	Test Performance, Groups I-VI	16
6	Test Performance on Printed Tests Plotted by the Average Time Required for Each Procedure	20
7	Plot of Test Performance for Subgroups Ia and Ib	21
A-1	Aircraft Recognition Features	30
A-2	Arrangement of Multi-Image Cards	31

Comparison and Evaluation of Printed Programs for Aircraft Recognition

INTRODUCTION

Military Problem

Visual recognition is used by all crews of forward area air defense weapons for aircraft identification. The development of new air defense weapon systems in recent years has stimulated increased interest in aircraft recognition skills.

Aircraft recognition training in the U.S. military services has traditionally consisted of group instruction using projected slide images. However, such an approach needs to be supplemented in many Army units because they also need training materials that can be used for self study or with very small groups (or individual trainees) on a highly flexible training schedule. These needs could be met with training materials that use printed images.

To ensure training effectiveness, such materials should include clear, simple directions that have been tested for effectiveness with the materials. Unless the materials are used in essentially the same way as when they were tested, there is no reason to expect them to be effective. Also, there is a need for testing procedures which can be used as easily as the printed training materials, in order to ensure that every man has attained an adequate level of performance.

Generally, a desirable level of recognition accuracy would be 90% to 99% (1, p.5). For example, the current Army Field Manual on Visual Aircraft Recognition specifies a class average of 90% recognition accuracy (2, pp. 5-6).

Research Problem

The objective of the research was to develop and evaluate an effective and efficient prototype of a printed, self-instructional aircraft recognition training program. This training program must incorporate training management procedures (including the methods for use and testing) which provide control over the product to be achieved through training.

Several kinds of prototype printed programs were developed and compared experimentally. The comparative evaluation was conducted not only to determine which program was best, but also to assess the performance level produced by the best program and the amount of training time required.

General Characteristics of Effective Programs

Since only a few prototype programs could be evaluated experimentally, the programs compared should be those which were most promising. Research and past experience with aircraft recognition training methods provided guidance on several general characteristics of effective aircraft recognition training programs.

The first formal training program for aircraft recognition was developed in England early in World War II, and Gibson (3) has noted that psychological theory at that time could provide no clear guide (see review by Vicory, 4). The method the British developed is known as the WEFT system (Wings Engine Fuselage and Tail) and consists of memorizing a series of details about these four major aircraft components, described verbally. The WEFT system typically used silhouettes of only three plan views (belly, head-on, side) for the analysis of features training. In 1941 the WEFT system was also adopted by the U.S. Navy and Army Air Corps.

Preceding page blank

The WEFT system has been criticized on several grounds (3): the heavily verbal character of the learning, overemphasis on features that could be named, and lack of systematic selection of those cues which are useful for actually distinguishing the aircraft from each other.

In 1942, Dr. Renshaw of Ohio State University introduced a radically different training method, using short (tachistoscopic) exposures of stimuli, under the rationale that such short exposures emphasized the *whole-image* concept of training rather than an *image-analysis* concept which, presumably, might induce a person to respond erroneously to a small part of the total form. The Renshaw system can produce a high level of identification accuracy with 1/75 second exposures, which was quite impressive. In 1942 the U.S. Navy adopted the Renshaw system, and in 1943 the Army Air Corps accepted a modified version of the system.

Exposure Interval. There seems to be no operational requirement to identify aircraft in less than a second, and Gibson (3) conducted an experiment which indicated no advantage for training on very short exposures (1/50 second), unless the testing also uses such short exposures. Thus, an effective training program would not necessarily include tachistoscopic image exposure—a point that is critical to the feasibility of an effective printed program.

Image Analysis and Distinctive Features. Gibson also found better performance resulting from emphasis upon the aircraft features early in training, especially emphasis upon those features which distinguished similar planes from each other (3, p. 131). The more general research on concept identification has indicated that performance is degraded by irrelevant attributes (5, 6), but that the disadvantage can be ameliorated by pretraining on the relevant attributes (7, 8, 9). Also, more discriminating attributes produce better identification of concepts (5). It may be inferred that an effective aircraft recognition program would use image analysis and would emphasize at first those distinctive features which are most useful for the discriminations required.

Since World War II, U.S. instructors have typically used some combination of WEFT and tachistoscopic procedures, often supplemented by practice with flash cards (which were first developed during World War II). The flash cards usually have shown the three plan views, and were typically used outside of formal class instruction.

About 15 years ago, the British developed another method, the Sargeant System, named after its originator, Charles Sargeant, who was editor of the *Joint Services Recognition Journal*. In the Sargeant System, an aircraft is first defined in a key which incorporates a verbal description of its presumably distinctive features, along with several named photographs of the aircraft. After studying the key, the student compares it with several test photos as he attempts to identify these photos. The Sargeant System is somewhat similar to the WEFT system in that it uses image-analysis techniques, but has certain apparent improvements in being somewhat more selective in the features discussed, and in providing more actual practice in identifying aircraft.

Simultaneous Comparisons. Another important feature of the Sargeant System is that it provides for *simultaneous comparison* of images. Gavurin (10) found that when the aircraft were displayed simultaneously (all at one time) the training was significantly better than when they were displayed successively (one at a time), even though in the criterion test the aircraft were displayed successively. Apparently there is some advantage in a learner's being able to make comparisons. Recently Perrin (11) sketched a theory of multiple-image communication, stressing the optimal "information density" for various purposes, and the organization of images to induce particular concepts. Studies of concept identification have also indicated the advantages of simultaneous presentation (12, 13), especially with greater numbers of simultaneously available instances and with more complex problems (14, 15, 16, 10). Thus, when trying to form the visual concept of an aircraft, it would be better to see several views simultaneously, rather than seeing them successively.

Selection of Views. One shortcoming of most of the aircraft recognition training methods in current use is that they use mostly "targets of opportunity" (readily available imagery) rather than views that have been systematically selected for their training value. HumRRO studies (17) have shown that the views a person sees during training markedly affect the views from which he can recognize an aircraft, and that a ground observer should be trained on about nine views of each aircraft if he is to have reasonably "flat" generalization curves (i.e., equal proficiency regardless of view). Also, targets of opportunity often have backgrounds so distinctive that the trainees may well learn to identify the background in the slide rather than the aircraft features. Aircraft images currently being used for training have another disadvantage: they often are so large as to present distinctive details which are very unlikely to be available as cues to the ground observer at tactically realistic distances.

It may be inferred that an effective program for aircraft recognition should use several views of the aircraft as it might be seen by a ground observer, that distinctive background should be eliminated, and that rather small images should be used to eliminate distinctive details. (These study results have been used as the basis for the GOAR kit requirement prepared by the USAADS.)

Prompting. Studies in response prompting (5) suggest certain other generally effective training methods which may be used in training aircraft recognition. Early in training it is probably better to prompt the student rather heavily (i.e., tell him the answers) rather than letting him guess wildly (18, 19, 20, 21, 22). Later, when he is practicing identification of aircraft, it is probably best to let him continue to see the aircraft image as the name is given to him (Stimulus-response overlap, 23, 24).

Previous Applications of the Principles

Classroom Method Using 35mm Slides. A method of training aircraft recognition in the classroom (1) was developed by HumRRO in 1965 to remedy many of the deficiencies of previous techniques. Historically, that classroom method was a direct antecedent of many of the printed imagery techniques compared in the present report (related, in particular, to those techniques which appeared to be most effective in the present evaluation).

The 1965 HumRRO method used an image-analysis approach with selected features and simplified nomenclature, and did *not* use tachistoscopic exposure of images. Simultaneous presentation of 35mm projected images was used for paired comparisons among aircraft which were similar at any particular view. Also, each student was issued a sheet with the three plan view silhouettes for each aircraft, along with the names, so the student could make other comparisons during training as he wished. Frequent testing was used to assess student progress, to increase motivation, and to assign remedial training when performance did not meet the desired standards. The slide images were 10 views of each aircraft, representative of the views a ground observer might see, without distinctive backgrounds, and of reasonably small image size.

The HumRRO classroom method was tried out with 16 aircraft for 16 class periods (50 minutes per period) and the class average was 95% at the end of the 16th session. This is impressive performance, well within the desired range of performance (90% to 99%, 1, p. 5) although the training time per aircraft might be considered somewhat high for many applications.

Training Method Using Printed Imagery. In preparation for the research described in the present report, a preliminary study was conducted. It involved the development and application of a series of aircraft recognition procedures using printed imagery, including: (a) Multi-Image Cards (each card pictured an aircraft from five views and described its significant features), (b) Paired-Comparison Cards (each card pictured two or three aircraft which appear to be similar from that view), and (c) Flash Cards (each card had one

view of an aircraft and its name). These practice materials, covering six aircraft, are described in greater detail in the next section. Practice procedures in the exploratory study were also similar to those described in the next section, except no printed form of test was given, and the instructions were relatively informal. Achievement was tested by the GOAR (Ground Observer Aircraft Recognition) slide test. The GOAR kit (based upon previous HumRRO research, 1) contains 10 views of each aircraft, and is being adopted by the Army for classroom training.

Experimental classes were conducted on 9 and 10 December 1969. One group (N=11) practiced with all three kinds of training materials, and averaged 79.1% on the GOAR test (three men were eliminated for apparent inattention). Another group (N=16) practiced with only the Multi-Image Cards and Flash Cards, and averaged 70.9% on the GOAR. The difference in achievement suggests that inclusion of the Paired-Comparison phase leads to higher achievement, although the difference was not statistically significant. The training and testing for each group was completed easily in one morning with considerable time to spare.

The results of the preliminary study indicated that a printed program could produce a rather high level of recognition accuracy in a reasonable amount of time, although there was considerable room for improvement. It was decided that further research (as described in this report) should evaluate not only the procedures used in the preliminary study, but also several alternative training methods using printed imagery. The alternative procedures were suggested by various researchers at HumRRO Division No. 5 who were especially interested in the problem of aircraft recognition training.

METHOD

Materials

In 1969-70 several aircraft recognition training procedures using printed images were assembled into experimental programs for comparative evaluation. All programs covered the same six aircraft (Skyhawk, Phantom, Freedom Fighter, Flashlight, Fishbed, Fishpot) as were used in previous view-generalization studies (17). The aircraft represented various levels of similarity and therefore are likely to be fairly representative of the difficulty level which would be encountered generally. With each of the component procedures, standard printed directions (Appendix A) were given out and the directions were read aloud as the students read them silently. All printed imagery (except in the Sargeant System) was the same rather small size (approximately 1" wing span). The component training procedures are described below.

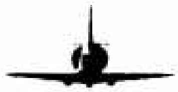
(1) Multi-Image Cards (MIC). Each of these cards pictured five different views of one of the aircraft, along with a brief verbal description of its most distinctive features. (See sample in Figure 1.) The five views, designated by heading and climb, were: 0-0, 90-0, 0-90, 190-15, and 340-15. The subjects would first study each card to get a general concept of each aircraft, then spread the cards so that all the aircraft could be compared at each view.

The training with Multi-Image Cards was designed to be given early in training. Therefore, distinctive features were pointed out early in this phase, and guessing was not encouraged. In the latter part of this procedure, comparisons were made across aircraft, and the practice conditions began to resemble testing conditions more closely. The multi-image cards were designed to include all the information that a student could readily use at this stage of learning, yet organized so as to facilitate desired comparisons and minimize search time for any particular piece of information.

(2) Paired-Comparison Cards. Each of these cards pictured two (or occasionally three) aircraft from the same view. (See sample in Figure 2.) Occasionally three aircraft

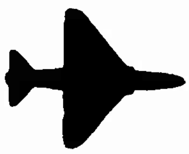
Sample Multi-Image Card

A-4 SKYHAWK A-4



Low mounted wings.

Air intakes high on sides of body, well above wing base.



Wings are delta shaped but rounded at the tips.

Mid-mounted tail flat.

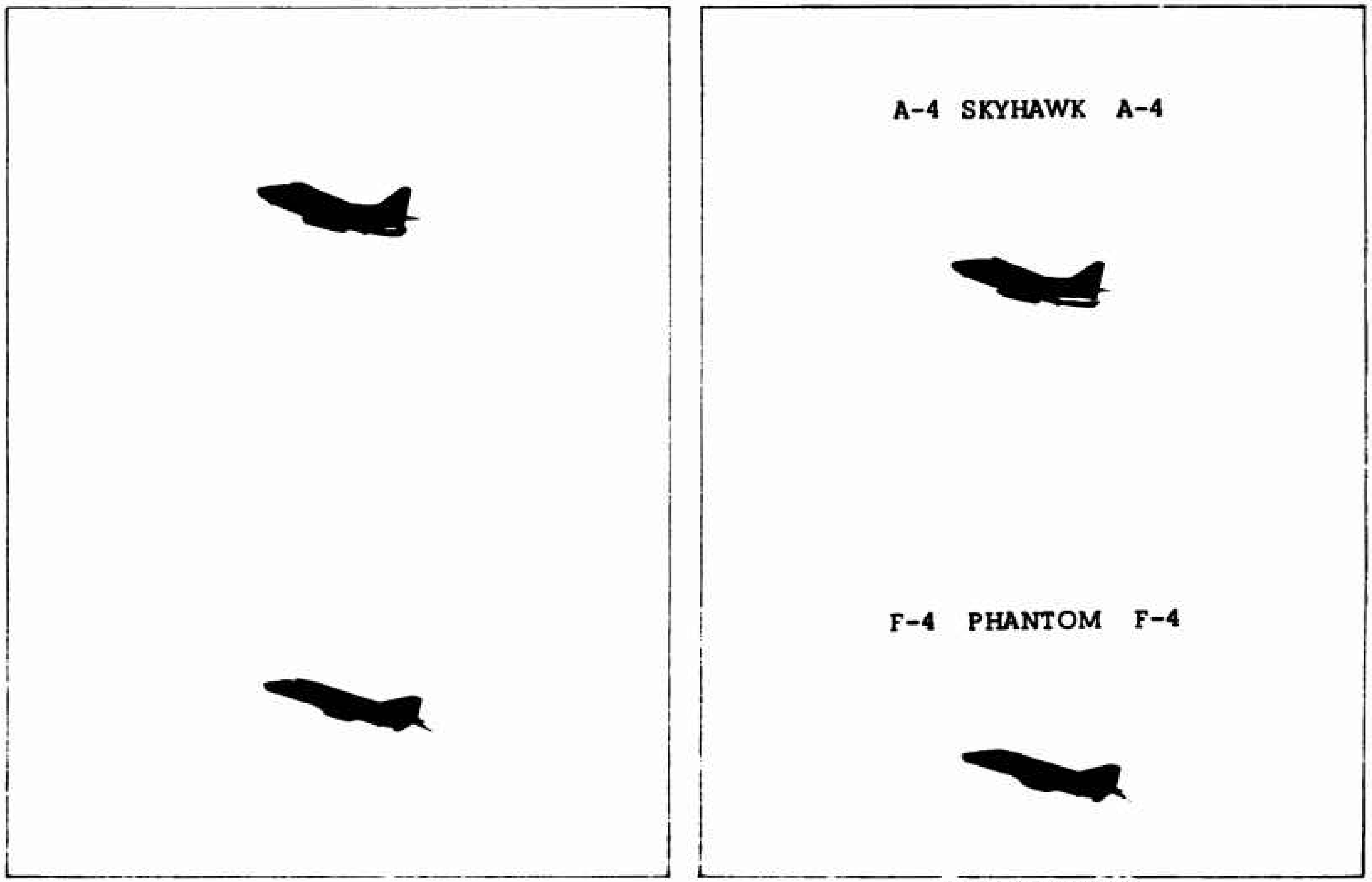
Single exhaust in the tail.

Stubby pointed nose.

A-4 SKYHAWK A-4

Figure 1

Sample Paired-Comparison Card



(reverse side)

Figure 2

would appear on one card when all three were similar at that view. The aircraft names were under their pictures, and on the reverse side were the same pictures without the names. The students were eventually to be tested on 10 views of each aircraft (see "Experimental Criteria").

The pairs (or triplets) of images were chosen from the total test set on the basis of which aircraft were confusable at that particular view. Confusability was first rated by two staff members, and the agreement was so high that more elaborate rating procedures were not employed.

Forty of the 60 test images were involved in the paired comparisons. With these Paired-Comparison Cards, the students first studied the images along with the aircraft names, then turned over the deck to practice naming the aircraft.

The paired-comparison drill was designed to build upon the practice provided with the Multi-Image Cards, providing for comparisons underlying the most difficult discriminations which a student would have to make.

(3) Flash Card Drill. Each of these cards pictured an aircraft on one side, and the same picture with the aircraft name on the other side (Figure 3). There was one card for each aircraft for each view to be tested (60 cards in all). Each student held his deck of cards with the names away from him and tried to name each picture as it appeared. After attempting to name an aircraft, the student would turn over the card to reveal the right answer Students were instructed to follow a "drop-out" procedure; that is, after checking his answer, the student would drop the card out of the deck if he had answered correctly, but return it to the deck if he had not been able to name the aircraft. Thus, an item would keep reappearing until the student answered correctly. After all the cards were thus eliminated, the whole drop-out procedure was repeated.

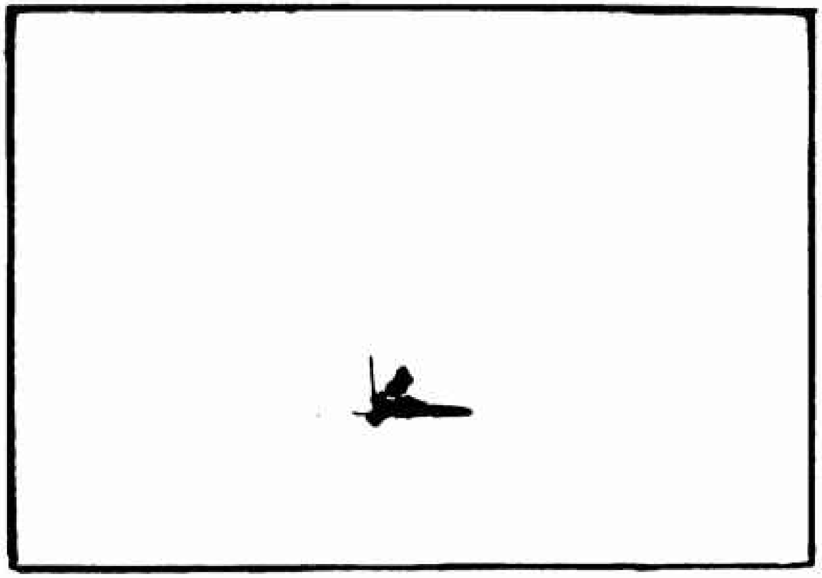
The flash-card procedure was designed for a rather advanced difficulty level, since it requires performance under circumstances very much like the test conditions. (The underlying factors have been discussed in the Introduction under "Prompting.") However, any undesirable effects of guessing are apt to be ameliorated somewhat by the fact that the image is printed on *both* sides of the card, thus providing for temporal overlap between stimulus and response terms (the same is true of the paired-comparison drill already described).

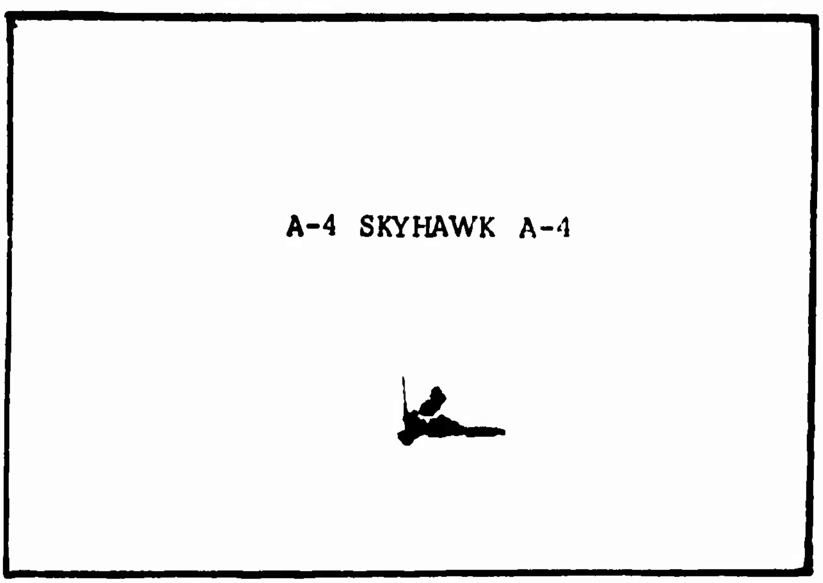
(4) Sargeant. With this procedure the students first studied a key (Book I, Figure 4A), which shows a few views of each aircraft along with a verbal description of the most distinctive features for each aircraft. The verbal descriptions are exactly the same as those used on the Multi-Image Cards. Next, the students attempted to identify the aircraft in Book II (Figure 4B), referring back to Book I as needed, and checking their responses with an answer key. The aircraft views in Book II which the subjects were to identify included all of the 60 views on which they were later tested. Book II had 120 items, each view twice, various size images.

(5) Sorting. Each subject was given a stack of 60 cards (one card for each view of each aircraft) and a sorting board. The sorting board had six spaces, one for each aircraft. Above each space was a verbal description of the distinguishing features of one of the aircraft (the same description as on the Multi-Image Cards). The names of the aircraft were not visible on this trial, so the sorting had to be done on the basis of the printed cues. Each student tried to sort the cards into six stacks, one for each aircraft, using the verbal cues to help define the stacks.

After finishing the first sorting, the student would collect the cards and put them back in the original order (the cards were numbered 1 through 60 on the back side). The student would repeat the sorting a second time, but a flap was unfolded from the back of the board revealing the names of the aircraft, so that the student could see both the cues and the aircraft names on this trial. Finally, the student re-ordered the cards and sorted them a third time, but this time with only the names of the aircraft

Sample Flash Card



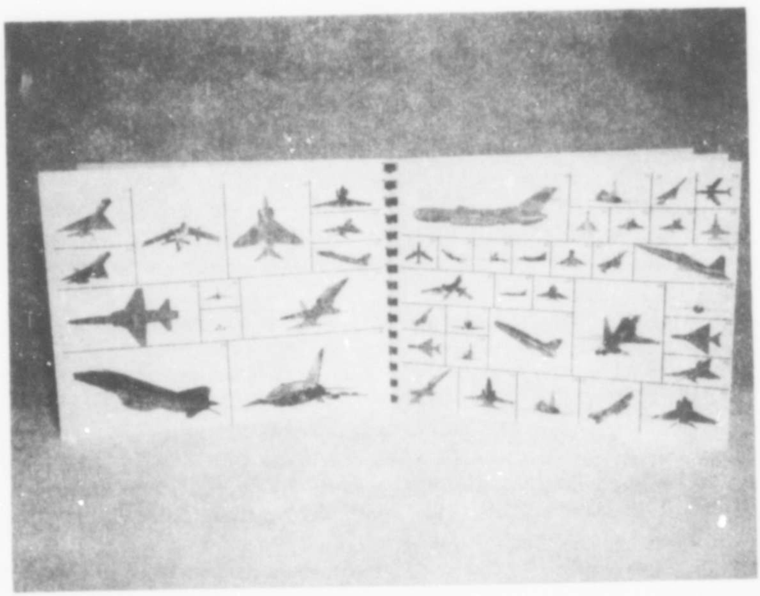


(reverse side)

Figure 3

Sargeant Materials

A–Sargeant Key (Book I)



B–Sargeant Problems (Book II)
Figure 4

visible (to cover the list of cues, the flap was folded so as to cover the upper margin of the board, and the aircraft names were visible on the top of the flap).

It should be noted that the students could be responding on the basis of position alone and ignoring the aircraft names, even on the third trial, because the board position for each aircraft was constant on all trials. Before the first test was given, the students would be urged to take a few minutes to memorize the names of the aircraft. The students would be instructed to try to write down all the names from memory. The instructional methods for teaching the names were brief and informal, and probably varied somewhat in effectiveness from one class to another.

The rationale underlying the sorting method was to induce the student to form his own perceptual concepts for each of the aircraft, while not requiring learning the aircraft names at first. Presumably, in making the multiple comparisons to generate perceptual concepts for each stack, the student would be likely to attend to a rich variety of cues. The absence of names, at first, should further encourage the subject to notice the various features of each aircraft.

(6) Sorting Game. This procedure was a competitive game based upon the sorting procedure (as described above). The first "hand" would be conducted under the sorting procedure, then the game element would be introduced. On the second hand, each man would choose an opponent to play against, and they would both play on a common board. The opponents would have the cards in the same order, and turn over each card at the same time. The object of the game was to place each card in the correct space before the opponent. Bonus points were given for catching the opponent's mistakes. (The standard directions are in Appendix A.) The game was devised to increase student motivation through competition.

Experimental Criteria

The ground observer aircraft recognition (GOAR) kit imagery is perhaps the most readily acceptable criterion of recognition performance which can be administered in the classroom. At the end of training, a GOAR-type slide test was administered to all experimental training groups. The locally produced GOAR slides were, for the most part, those used in the preliminary study, but several slides were replaced because of somewhat dubious image quality.

The slides were exposed at eight-second intervals. First there were eight practice slides, on which the answers were given aloud, followed by the test slides. It was felt that students might require more than the eight practice slides to become accommodated to the test situation, because in the preliminary study students made more errors on the first half of the test than on the last half. Therefore, the first 30 "test" slides were not scored; another 60 slides were presented (all views in random order) and scored as the criterion test.

However, it was also desirable to measure progress of the various groups at several points during their training, in order to assess the individual effects of particular procedures, and repeated administration of a slide test is somewhat cumbersome. Therefore, a printed test, using the same imagery as the GOAR, was prepared. The test booklets were sets of cards like the flash cards, except that they were printed on only one side, without the names, and the 60 different items were bound together with a ring binder. Each test booklet used a different random order, and each student was given a different booklet on successive tests. The pages were numbered (item number) and the answers were recorded on separate answer sheets.)

In scoring the answer sheets, either the code name (e.g., Skyhawk), its abbreviation (e.g., SH), or its number designation (A4) was considered acceptable. The guideline was that any response was acceptable if it distinguished the aircraft from the others in the set being learned.

In evaluating the various treatments, the time required for the training was also considered. On each component procedure and on each test, a research assistant would mark the beginning of the period on an event recorder, and also record when each student finished. This would yield an approximate distribution of times for each of the subperiods.

Experimental Groups

The combinations of procedures which were tried represented various feasible training programs. Since a rather high level of criterion performance was desired, it was decided that for all groups enough training procedures should be provided to fully utilize the time available (one-half a working day, minus time for transporting and assembling the trainees). In the preliminary study the training left considerable room for improvement, and past research (1) indicated that the desired performance level required a rather long time in training. If fewer training procedures were required to bring the subjects to a satisfactory level of performance, this fact would be apparent from their performance on the printed form of the test, which was administered repeatedly (to all groups except VII). If sometimes there was insufficient time to finish all the planned procedures and the GOAR slide test, the training would be terminated soon enough so that the GOAR could be administered.

Relatively minor variations in training procedure were designated as (a) and (b) subgroups, so that these groups might be pooled if no substantial differences resulted. The training programs administered to the various groups are outlined in Table 1 (the component procedures have been described in the previous section).

Table 1

Training Programs Administered to the Various Groups

Group		Training Program[a]										N
I	(a)	MIC	x	PC	x	FC	x	Sarg	x	G		14
	(b)	MIC	x	FC	x	PC	x	Sarg	x	G		14
II	(a)	Sort	x	PC	x	FC	x			G		9
	(b)	Sort	x	FC	x	PC	x			G		10
III	(a)	Sort	x	PC	x	Sarg (only started)				G		7
	(b)	Sort	x	PC	x	S. game (only started)				G		9
IV	(a)	S. Game	x	PC	x	Sarg				G		12
	(b)	S. Game	x	Sarg	x	PC				G		12
V		FC	x	FC	x	Sort	x			G		15
VI		Sarg	x	MIC	x	S. game x (abbreviated)				G		15
VII		Potpourri (all procedures sampled)					x			G		18

[a] x = printed form of test; G = GOAR slide test
MIC = Multi-Image Cards; Sarg = Sargeant System
PC = Paired-Comparison Cards; Sort = Sorting procedure
FC = Flash Cards; S. game = Sorting game

With Group III, the third phase (Sargeant or Sorting Game) was cut short because there would not have been enough time for the GOAR test if this phase had continued. There was only time for reading the directions and looking over the materials for about two or three minutes. With Group IV there was sufficient time for the third phase training, but insufficient time for the third printed test. With Group VI, the third procedure (Sorting Game) was somewhat abbreviated by omitting the third hand, leaving one sorting hand and only one hand of the game.

The training program for Group VII consisted of a sampling of all the procedures. The students were instructed not to spend much time on any one procedure, and time was called in each phase when several men had not yet finished. The program was rather loosely structured, compared with the other programs. The intention was to give all these students a sample of all procedures, then a test, followed by further practice on whatever materials they chose, but there was not time enough for the last practice stage. The order of the procedures for Group VII (and the time spent on each) were as follows: Multi-Image Cards (21 min.), Sargeant (16 min.), Paired-Comparison Cards (13 min.), Sorting Procedure (one hand, 32 min.), Sorting Game (one hand, 22 min.), Flash Cards (11 min.).

Students

The experimental students were 135 enlisted men from Fort Bliss, distributed among the groups as shown in Table 1. None had previous formal training in aircraft identification, all had a GT score of 90 or better,[1] and all had 20/20 vision (sometimes corrected with glasses). Each treatment group was a sample from various organizations on post in roughly equivalent proportions (except for Group VII, which came as an intact group). The motivation of the groups was expected to be somewhat lower than that of men who might ordinarily be taking aircraft identification, since none of the experimental students needed aircraft identification for their military occupational specialty.

Students who scored at a chance level or below on their first test (a score of 10 out of 60) were considered to have inadequate motivation and therefore were dropped from the analysis (9 persons were thus eliminated).[2] Two additional students were eliminated because of failure to complete their training (one was called out on an emergency, one refused to continue).

Procedure

The students were trained in groups of 9 to 20 men per class; the number varied depending upon the availability of men on that day at the units supplying the troops. Three instructors shared the administrative duties (passing out materials, recording time, etc.). Thirteen half-day classes were conducted during the period 16-27 March 1970. The three-man instructional team sometimes conducted two training programs concurrently in adjacent classrooms; generally, it would be the (a) and (b) divisions of the major programs that were conducted concurrently. The subgroups would be started together, and separated into different rooms when their procedures differed. It usually required two classes to constitute an experimental group.

[1] GT—General Technical Aptitude Area tests from the Army Classification Battery for classifying enlisted personnel.

[2] There were four or five subjects in various groups who scored very low on their first test and even worse later, after some training. Typical of these men was a final test with more than half the spaces blank, not even a guess, and some apparently random aircraft names in the other spaces, with a final score of four or five out of 60. Retaining data from these men would have seriously compromised the sensitivity of the experiment. The criterion used also eliminated some men who later did well but was adopted to maintain an objective criterion for eliminating subjects.

After each class was assembled, an instructor made brief introductory remarks on the importance of aircraft recognition and what was expected of the students. The men were told that they would be in the class one-half day, that they would be expected to learn to identify six aircraft, and that the purpose of their participation was to evaluate the effectiveness of the various experimental training materials.

After the introductory remarks, the training procedures were administered as indicated in Table 1. Those men who finished a procedure or the test before the rest of the class would wait, and all would begin the next procedure together.

RESULTS

Preliminary *t* tests were run between the (a) and (b) subgroups on the last printed test and on the GOAR for each of Groups I, II, III, and IV. None of these was significant at the .05 level so the subgroups were pooled for further analyses.

The mean test scores of the various experimental groups on all tests are given in Table 2. These values are plotted in Figure 5, except for Group VII, whose first test came at the end of training, and would therefore be somewhat misleading in the plot.

Table 2

Mean Test Scores

Group	Printed Tests				GOAR
	1	2	3	4	
I	78.5	90.0	94.4	94.6	87.2
II	72.1	80.3	85.4		72.8
III	69.9	84.6			73.7
IV	78.9	87.5			81.9
V	75.6	87.8	86.7		81.8
VI	72.6	85.1	86.1		78.0
VII	86.0				79.9

The best criterion of recognition performance at the end of training would be either the last printed test or the GOAR. Use of the last printed test might be criticized because some groups have taken this form of test more often than other groups, so their scores may reflect a greater ability to take tests, which might not transfer to recognition of actual aircraft. This argument is somewhat mitigated because all groups (except VII) have more than one such test experience, and the effect of the fourth test on Group I seems to be minimal, since its curve appears to approach asymptote on test 3.

Also, the practice materials for all groups very closely resemble the test material, and the tests were designed to cover rather thoroughly the views of relevance for ground observers, so that "test learning" would not represent merely a small sample of the desired performance. The same arguments would apply somewhat to the GOAR test because it was very similar to the printed tests even though the medium was different. Note, however, the special case of Group IV, which had one phase of instruction between the last printed test and GOAR so that the last printed test does not reflect learning in the last phase of practice. But Group IV also took more practice time than other groups with which it was compared, so it is doubtful whether the last practice phase should be included (see subsequent discussion of time required).

Test Performance, Groups I-VI

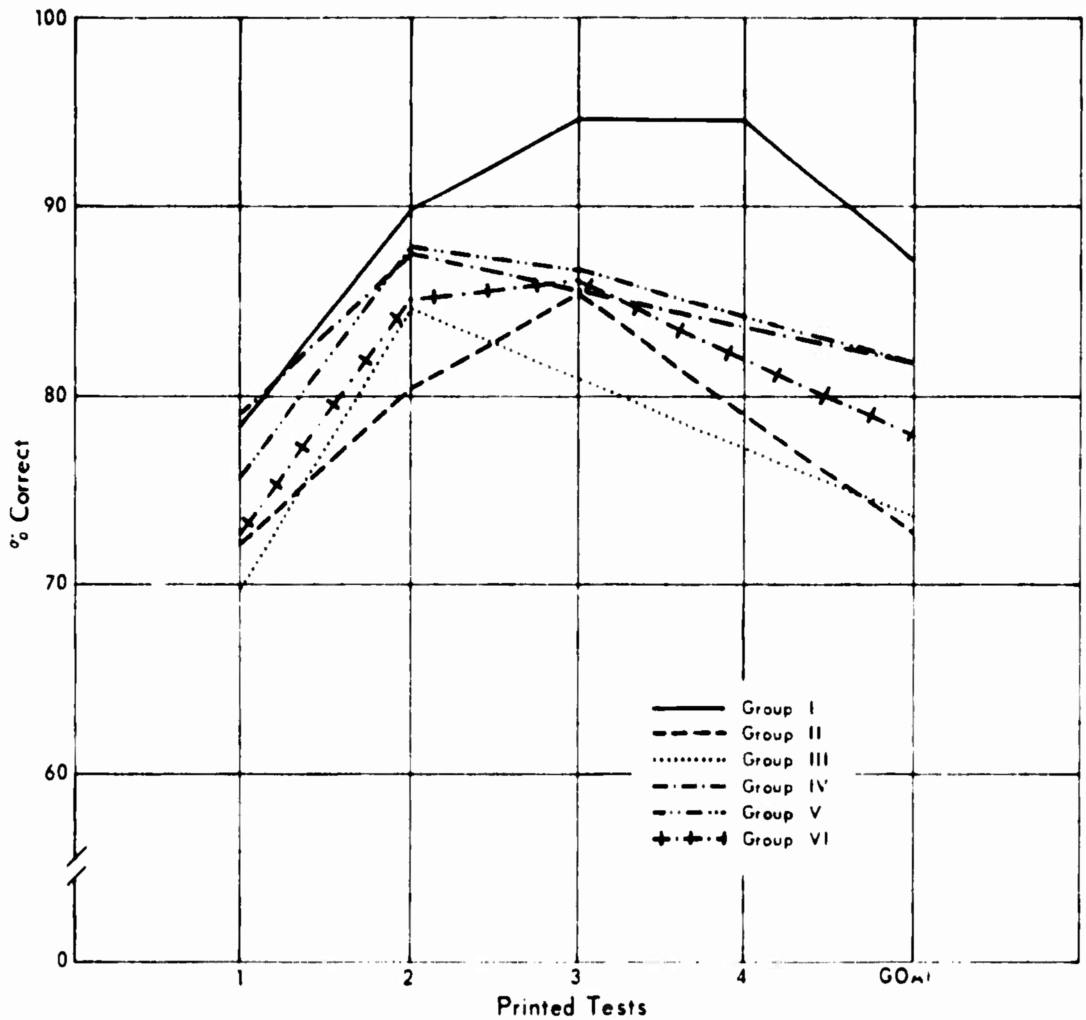


Figure 5

Both the last printed test and GOAR yield the same pattern of results on the critical issues, differing primarily in level of significance on particular comparisons, so there seems no need to decide sharply between them. Therefore, both criteria will be considered in evaluating the results.

A preliminary test of homogeneity of variance indicated that the variances for the groups were significantly different, both for the GOAR (Fmax = 4.1, $p < .05$) and for the last printed test (Fmax = 6.38, $p < .01$). As a result, use of conventional analysis of variance was not justified. Inspection of the group variances revealed that Group I had an especially small variance on both tests; the constricted variance was apparently caused by a "ceiling" effect (i.e., Group I means were so close to a perfect score that no score can be far above the mean). Group I, on the last printed test, made less than half as many errors as the next closest group. (The distributions are given in Table 3.)

Table 3

Distribution of Scores After Training

% Correct	Last Printed Test							GOAR Test						
	Training Method							Training Method						
	I	II	III	IV	V	VI	VII	I	II	III	IV	V	VI	VII
100	13	6	2	7	4	3	2	1			2	1		
98.3	5	3			2	2	5	5			1	1		
96.7	4	3		3	2	2		2	3		4		3	1
95.0	1	1	2	2	1		2	2	3	1	1	2		2
93.3			1	2		1	2	1		2		2	1	
91.7		1	3	1		1		3				1		1
90.0		1	1				1	2	1			2		
88.3				1	1			1		1	1			3
86.7				1				2			2		1	1
85.0			1			3		2	1		2			1
83.3	1						1			1	2		1	2
81.7			1	2	1			2						
80.0			1					1	2		2		1	
78.3				1	1		1		3	1		1	1	2
76.7			1	1			1			2	1			
75.0	1										1		2	1
73.3	1													1
71.7	2							1			1	1	1	
70.0				1						2			1	
68.3			1							1				
66.7			1								1			
65.0					1		1			1		1		
63.3										2			1	
61.7						2		1					1	1
60.0							1				1			1
50-59					1			2	2	1		3		
40-49		3	1	1	1				1	1	1		1	
30-39		1		1		1			1		1			1
20-29							1		1					
10-19									1					
n_j =	28	19	16	24	15	15	18	28	19	16	24	15	15	18

No particular transformation of the data seemed particularly justified, so nonparametric tests were run. The variation among groups approaches statistical significance, both for the last printed test (Kruskal-Wallis H = 12.17, $p < .10$) and for the GOAR (H = 12.12, $p < .10$). Both of these "H" values are just short of the critical value for the .05 level of significance, so they seemed sufficiently suggestive to justify further tests among particular groups. However, the further tests should be interpreted with an extra measure of caution because of the marginal significance of the overall test. Also, there is no nonparametric test which adjusts the probability level for the number of comparisons being made (as would the Newman-Keuls test) so the probability levels should be interpreted with extra caution.

The results of the tests of statistical significance are given in Table 4. (Since the tabled values are one tailed tests, $p < .02$ will be employed.) Group I was significantly superior to each of the other groups except Group V on either GOAR or the last printed test ($p < .02$), and all of the differences somewhat favored Group I. It seems only reasonable to conclude that the training received by Group I is superior to most of the other training methods, and is not likely to be appreciably worse than any of them.[1]

Table 4

Significance Tests[a] of Differences Between Group I and Other Groups

Groups	Last Printed Test	GOAR Test
I:		
vs II	$p < .10$	$p < .02$
vs III	$p < .001$	$p < .001$
vs IV	$p < .02$	$p < .15$
vs V	$p < .04$	$p < .25$
vs VI	$p < .02$	$p < .02$
vs VII	$p < .01$	$p < .02$

[a]Mann Whitney U One tailed tests

The first test is another reasonable point of testing for differences among groups, since most groups have had only one kind of procedure at that point. The preparatory test for homogeneity of variance indicated there was no statistically significant heterogeneity (Fmax = 2.15). The means on the first test are not significantly different.

The times taken for the various component procedures are given in Table 5. These times do not include reading of instructions, introductory remarks, or reassembling of materials. For time comparisons among groups, testing times have been included; if testing times were disregarded, it would have to be assumed that the tests do not affect the learning process, which seems untenable. Also, if time for testing were disregarded,

[1] Parametric tests were also conducted, even though the assumption of homogeneity of variance was not tenable. The overall F ratio was non-significant at the .05 level, as were all individual comparisons by the Newman Keuls range test, both on the last printed test and on the GOAR. When individual comparisons were conducted by the simple t test between groups (the parametric test comparable to "U" tests reported in Table 2), only one of the eight comparisons was significant at the .02 level. Siegel (25, p. 126) states that the "U" test has *greater* power than the t test for some distributions; these data apparently are one such instance.

Table 5

Time Required for the Component Procedures (in minutes)

		TEST 1		TEST 2		TEST 3		TEST 4
Group I	MIC		PC		FC		SARG	
Mean	24.6	16.2	7.8	8.2	9.8	7.6	11.4	6.5
Range	22-31	6-29	5-11	5-12	5-17	4-12	8-16	4-9
Group II	Sort		PC		FC			
Mean	35.6	11.3	8.1	8.2	8.3	7.8		
Range	16-54	5-17	5-11	5-13	5-11	5-12		
Group III	Sort		PC					
Mean	46.6	12.2	11.1	8.4				
Range	39-61	11-15	10-13	7-10				
Group IV	S. Game		PC		Sarg			
Mean	78.7	12.2	11.3	9.0	15.7			
Range	63-89	5-19	7-14	5-16	7-16			
Group V	FC		FC		Sort			
Mean	28.3	13.4	9.9	9.3	33.1	7.8		
Range	16-35	7-23	6-14	5-14	15-42	4-11		
Group VI	Sarg		MIC		S. Game			
Mean	35.1	14.5	13.6	10.2	35.4	6.6		
Range	27-50	9-22	8-19	6-15	27-47	5-8		
Group VII	Potpourri							
Mean	115	8.8						
Range	115	5-14						

the trend obtained would be exaggerated, and the estimated time required for the apparent best method would be unrealistically low.

These times show that Groups I, II, and III took distinctly less time than the other groups, and that Group I is the briefest of all if time is cumulated only through the third test, at which Group I appears to rise to the high point in their performance curve. Since Group I also had distinctly the highest performance scores, it would appear to be the most effective and efficient method.

In Figure 6 the average scores on printed tests are plotted according to the average time required, which gives clearer comparisons of progress over time than does Figure 5 Since Group I about reached its highest achievement score on Test 3 (95%), there is little justification for also administering the fourth procedure (Sargeant). The mean total time to reach the 95% performance is 72 minutes, including administration of the three tests.

Since the method received by Group I is the apparent best method, it would be well to examine more closely whether method I(a) or method I(b) appears better—that is, whether the Paired-Comparison procedure should be given before or after the Flash Card procedure. The preliminary t tests for differences between (a) and (b) subgroups were not

Test performance on Printed Tests Plotted by the Average Time Required for Each Procedure

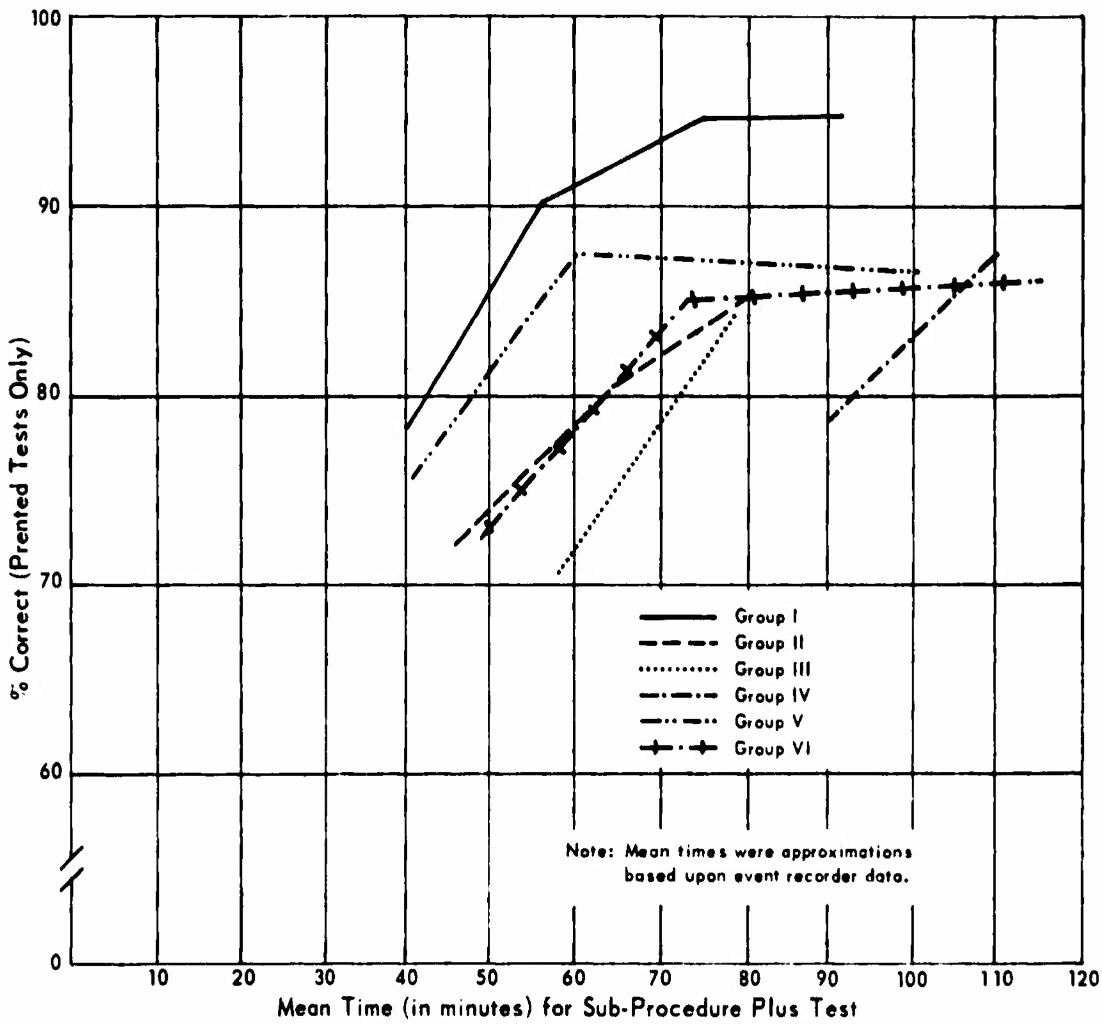


Figure 6

significant, but these tests were somewhat insensitive, designed to detect only differences so gross as to invalidate pooling procedures. The performance of the two subgroups is plotted in Figure 7.

The most sensitive test of such an order effect should be test 3, because this is the first test given after the differential treatment, and both groups have had all three procedures. It should also be noted that both subgroups received the same treatment up to test 1, and so should differ only by random assignment of the subjects to the two treatments. A test of this assumption indicated that it was tenable (differences between I(a) and I(b) on test 1 yields a t = .307, not significant). Performance on test 1 was therefore used as a covariate for an analysis of covariance on test 3.[1] The results indicated

[1] The use of the analysis of covariance has been criticized (e.g., 26), but the present analysis clearly seems to be a valid application because the groups did not receive *differential* treatment until after the covariate was measured.

Plot of Test Performance for Subgroups Ia and Ib

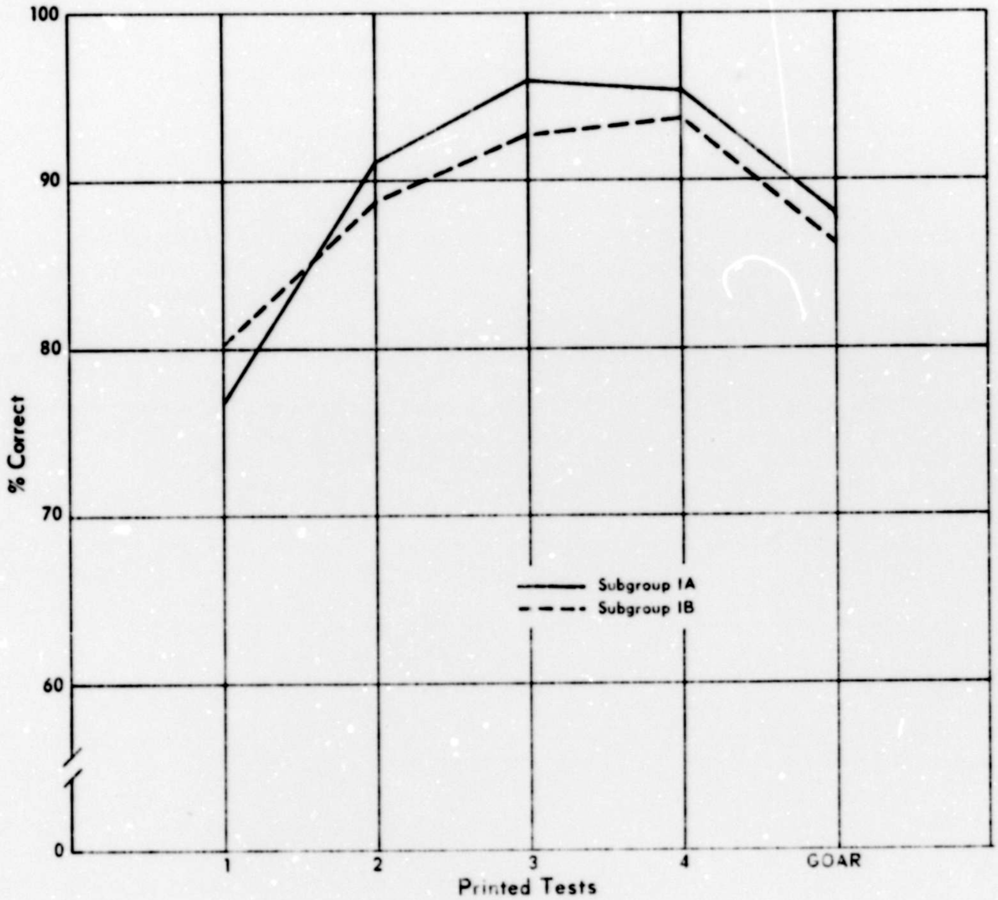


Figure 7

that Subgroup I(a) improves significantly more than I(b) ($p < .05$). It may be concluded that it is better to give the Paired-Comparison training before practice with the Flash Cards, rather than afterward.

For every subject a GT score was available, and these scores were correlated in Group I with performance on the last printed test ($r = .37$, $p < .10$) and on the GOAR ($r = .40$, $p < .05$). It may be concluded, therefore, that there is a modest but significant relationship between general ability (as measured by the GT) and performance on the aircraft recognition test.

DISCUSSION

The program received by Group I is the printed program to adopt, in terms of the practical decision. Group I scored the highest and took the least amount of time Its margin of superiority was sufficient, in most cases, to be significant statistically, either on the last printed test or on the GOAR slide test. It seems very unlikely that Program I would turn out, in the long run, to be appreciably worse than any of the other programs. Probably the last procedure (Sargeant) could be dropped, since there was no additional improvement with this procedure.

Other practical considerations also make Program I desirable. It is one of the easiest to administer, with rather straightforward instructions and no need for auxiliary equipment or procedures. It is easy to print, especially when the Sargeant phase is omitted, using only one size of images. Also, if the method is to be used in a particular theater of operations (where only a few aircraft would be likely), the various decks of cards could be sorted so as to include only those particular aircraft.

The Paired-Comparison Cards, however, may become somewhat more difficult to generate as the number of aircraft increases, because of the sharply increasing number of possible comparisons. The difficulties might be minimized by first separating the aircraft into rough similarity groups (e.g., one might assume that any helicopter would be distinctly different from any jet regardless of view) and then sorting them into confusable pairs or triplets.

The order of procedures for Subgroup I(a) (Multi-Image Cards, Paired Comparisons, Flash Cards) seems better than the order for Subgroup I(b) (with the Flash Cards coming before Paired Comparisons). The difference, though small, is statistically significant ($p <$ .05) by analysis of covariance. Even if the two groups had scored the same, Method (a) probably would have been chosen because it could be justified by a better rationale. The paired comparisons are designed for an intermediate stage of learning, after the general concepts have been formed, during which the student refines the concepts by practicing the most difficult discriminations he will have to make. The flash card phase, resembling more closely the final test, provides practice of the total skill with minimal guidance.

The printed tests were interspersed throughout the programs, and it is only prudent to keep them as part of the instructional procedure. If they were omitted, one could *not* assume that the same results would be obtained. The tests seemed to focus attention on the instructional objectives. The printed form of the tests is convenient to administer, and may be used as a quality control to assure that all students have reached the desired level of mastery at the end of training.

Although Group I apparently had the best training as a total method, it is possible that other component procedures might be as effective for particular phases. For example, the Sargeant System (first procedure for Group VI) produced a level of performance close to the Multi-Image Cards for Group I (not significantly worse) and required only a little more time. Actually, there is a distinct similarity in the operations the students perform under the two procedures. However, the Sargeant System, as applied here, could not be expected to be equivalent to the whole series of procedures administered to Group I.

Use of Flash Cards from the beginning (Group V) also produced rather high performance for the first two phases, without using much more time than Group I. The fact that the performance level regressed with the Sorting procedure does not necessarily reflect on the efficiency of the Flash Card procedure. However, use of this procedure initially has not been demonstrated to produce eventual high level skill, and other experiments (e.g., 3, p. 131) have shown the advantages of emphasizing distinguishing features early in training.

The Sorting procedure produced a somewhat lower level of performance than the Multi-Image Cards (not statistically significant) and took a somewhat longer period of time, but part of the lower scores might be spurious. Although the subjects were instructed to memorize the names, the Sorting procedure did not ensure response learning per se, and a few of the subjects after the first test reported that sometimes they could recognize an aircraft but had forgotten the name. Such failure of response learning, however, should disappear on the second test. Also, a few of the subjects initially seemed to have difficulty even going through the procedure of Sorting. These subjects would grossly misplace many of the cards, and would arrive at concepts of each aircraft only very slowly, and the lack of knowledge of results would allow the confusion to continue for a considerable period of time. However, most of the subjects seemed to perform the Sorting rather easily from the beginning.

The Sorting, and especially the Sorting Game, seemed more cumbersome to administer than the other procedures. Most of the extra operations, such as getting the cards back into sequence between sorts, are not reflected in the times in Table 5. The Sorting Game, although apparently reasonably effective, took the longest of any of the introductory treatments. If Sorting or the Sorting Game were to be implemented, perhaps as a supplement to the other procedures, the procedures might be simplified.

It is desirable to have a variety of aircraft recognition training procedures available, because a high skill level seems to require extended practice of the task. General observation during the experiment indicated that students would go through a new procedure even though they had "finished" their practice on the previous method.

The main purpose of the experiment was to identify one or more convenient training procedures which would produce a rather high level of performance in a reasonable period of time. In this, it seems to have succeeded. The absolute performance level of Group I is creditable, especially in view of the time taken. Group I(a) actually averaged 96% on their third printed test, at an average training time of 71.3 minutes, including the three tests, or about 12 minutes per aircraft. Although the average time per aircraft might rise somewhat as more aircraft are learned, the training time to be expected would still be very reasonable. As a result of this study, the Department of Doctrine Development, Literature and Plans of the Air Defense School has prepared a limited number of aircraft recognition training materials based upon the apparent best method.

LITERATURE CITED
AND
APPENDIX

LITERATURE CITED

1. Whitmore, Paul G., Cox, John A., and Friel, Don J. *A Classroom Method of Training Aircraft Recognition*, HumRRO Technical Report 68-1, January 1968.
2. Department of the Army *Visual Aircraft Recognition*, Department of the Army Field Manual FM 44-30, March 1969.
3. Gibson, J.J. (Ed.) *Motion Picture Testing and Research*, Armed Forces Aviation Psychology Program Research Reports, Report No. 7, U.S. Government Printing Office, Washington, 1947.
4. Vicory, A.C. *A Brief History of Aircraft Identification Training*, HumRRO Professional Paper 27-68, August 1968.
5. Archer, E.J., Bourne, L.E., and Brown, F.G. "Concept Identification as a Function of Irrelevant Information and Instruction," *Journal of Experimental Psychology*, vol. 49, pp. 153-164, 1955.
6. Bourne, L.E., Jr., and Haygood, R.C. "The Role of Stimulus Redundancy in Concept Identification," *Journal of Experimental Psychology*, vol. 58, pp. 232-238, 1959.
7. Haygood, R.C., and Bourne, L.E., Jr. "Attribute- and Rule-Learning Aspects of Conceptual Behavior," *Psychological Review*, vol. 72, pp. 175-195, 1965.
8. Haygood, R.C., and Stevenson, M. "Effects of Number of Irrelevant Dimensions in Nonconjunctive Concept Learning," *Journal of Experimental Psychology*, vol. 74, pp. 302-304, 1967.
9. Looney, N.J., and Haygood, R.C. "Effects of Number of Relevant Dimensions in Disjunctive Concept Learning," *Journal of Experimental Psychology*, vol. 78, pp. 169-171, 1968.
10. Gavurin, E.I. *An Evaluation of Various Tachistoscopic and WEFT Techniques in Aircraft Recognition*, Technical Report NAVTRADEVCEN IH 40, U.S. Naval Training Devices Center, Port Washington, N.Y., 1965.
11. Perrin, Donald G. "A Theory of Multiple-Image Communication," *AV Communication Review*, vol. 17, no. 4, pp. 368-382, Winter 1969.
12. Cahill, H.E., and Hovland, C.I. "The Role of Memory in the Acquisition of Concepts," *Journal of Experimental Psychology*, vol. 59, pp. 137-144, 1960.
13. Hovland, C.I., and Weiss, W. "Transmission of Information Concerning Concepts Through Positive and Negative Instances," *Journal of Experimental Psychology*, vol. 45, pp. 175-182, 1953.
14. Bourne, L.E., Jr., Goldstein, S., and Link, W.E. "Concept Learning as a Function of Availability of Previously Presented Information," *Journal of Experimental Psychology*, vol. 67, pp. 439-448, 1964.
15. Kates, S.L., and Yudin, L. "Concept Attainment and Memory," *Journal of Educational Psychology*, vol. 55, pp. 103-109, 1964.

16. Pishkin, V., and Wolfgang, A. "Number and Type of Available Instances in Concept Learning," *Journal of Experimental Psychology*, vol. 69, pp. 5-8, 1965.

17. Whitmore, Paul G., Rankin, William C., Baldwin, Robert D., and Garcia, Sandra. "Studies of Aircraft Recognition Training," HumRRO Technical Report in preparation.

18. Kopstein, F.F., and Roshal, S.M. "Verbal Learning Efficiency as Influenced by the Manipulation of Representational Response Processes," in A. A. Lumsdaine (Ed.), *Student Response in Programmed Instruction: A Symposium*, National Academy of Sciences—National Research Council, Washington, 1961.

19. Hawker, J.R. "The Influence of Training Procedure and Other Task Variables in Paired Associate Learning," *Journal of Verbal Learning and Verbal Behavior*, vol. 3, pp. 70-76, 1964.

20. Hawker, J.R. "The Effects of Training Procedure Response Availability, and Response Meaningfulness in Multiple-Choice Paired-Associate Learning," *Psychonomic Science*, vol. 3, pp. 329-330, 1964.

21. Hawker, J.R. "The Effects of Training Procedure, Response Similarity, and Number of Response Alternatives in Multiple-Choice Paired-Associate Learning," *Psychonomic Science*, vol. 3, pp. 331-332, 1965.

22. Peterson, L.R., and Brewer, C.L. "Confirmation, Correction, and Contiguity," *Journal of Verbal Learning and Verbal Behavior*, vol. 1, pp. 365-371, 1963.

23. Sidley, N.A., Winograd, E., and Bedarf, E.W. "Stimulus-Identification Overlap in Learning to Identify Complex Sounds," *Journal of the Acoustical Society of America*, vol. 38, pp. 11-13, 1965.

24. Sidley, N.A., Winograd, E., and Bedarf, E.W. "Learning to Identify Complex Sounds: Prompting Versus Confirmation," *Journal of the Acoustical Society of America*, vol. 38, pp. 1050-1052, 1965.

25. Siegel, Sidney. *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company, Inc., New York, 1956.

26. Evans, Selby H., and Anastasio, Ernest J. "Misuse of Analysis of Covariance When Treatment Effect and Covariate Are Confounded," *Psychological Bulletin*, vol. 69, no. 4, pp. 225-234, April 1968.

Appendix A

STANDARD INSTRUCTIONS

Instructions: Multi-Image Cards

Stage 1: Studying multi-image cards. Study each card carefully, one at a time. Read each recognition feature and look for it in the pictures of the aircraft. Whenever you have any doubt about the meaning of a recognition feature, see Figure 1 on the next page for illustrated examples. When you feel familiar with one aircraft, go to the next card. When you feel familiar with all the aircraft, go to Stage 2 (no test is given here).

Stage 2: Drill with multi-image cards. Spread the multi-image cards as in Figure 2 so that you can compare all the aircraft with one another from each viewpoint. See if you can name each image (cover the names where they show). Practice with one view at a time, working in a row across the cards, naming each aircraft. When you come to a view you don't know, uncover the name at the top of the card so that the next time through you will be able to correctly name the image. If you are having a lot of trouble naming the aircraft, go back to Stage 1 and review each card again with the name uncovered before returning to Stage 2.

You will have a half hour to study with the multi-image cards using the procedures just described.

Paired Comparison Practice

Step 1, study. Take the paired comparison cards and place them with the name side up. Look at the first card. These aircraft have been grouped together because they are often confused from this angle, so you'll want practice in telling them apart. Notice the differences between these aircraft. Then continue on to the next card. Go through all the cards, making comparisons in each group.

Step 2, drill. When you have finished the last card, *turn over* the pack so that the names are on the bottom side of each card. Go through the pack one card at a time and try to name the aircraft in each group, then turn the card over to check your answer. When you finish this, repeat the whole procedure (both steps 1 and 2), until you feel that you can name the aircraft correctly most of the time.

Practice with Flash Cards

Hold the deck of flash cards so the aircraft names are away from you. Go through the deck one card at a time and identify each aircraft, checking your answers on the back. If you didn't get the right name, put that card in the back so that it will come up again. But if you do get it right, drop that card out of the deck by placing it on the table. Continue going through the deck until all the cards have been dropped out, meaning you have correctly named them all.

When you get all the cards correct, shuffle the deck and work through it again using the same procedure.

Aircraft Recognition Features

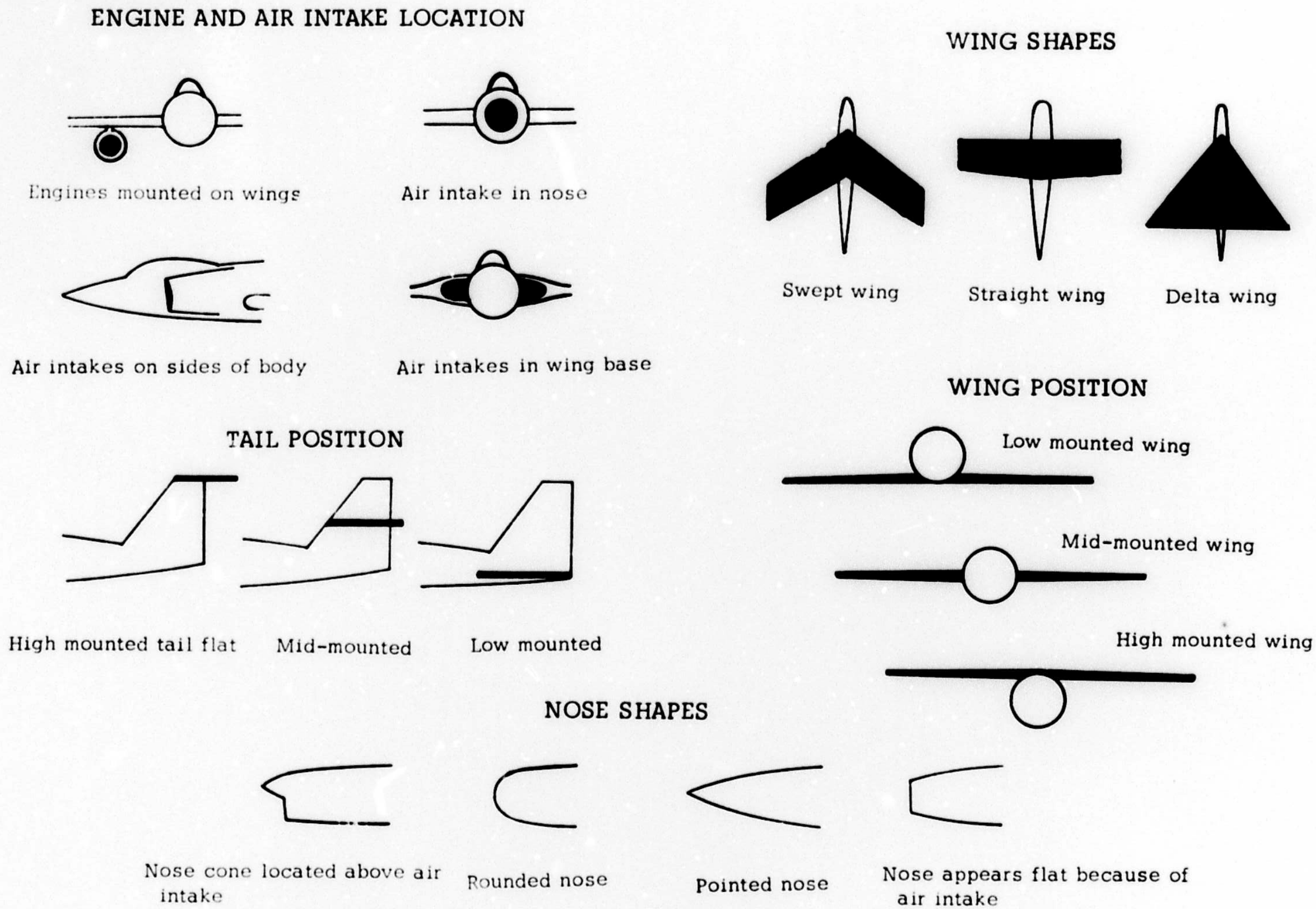


Figure A-1

Arrangement of Multi-Image Cards

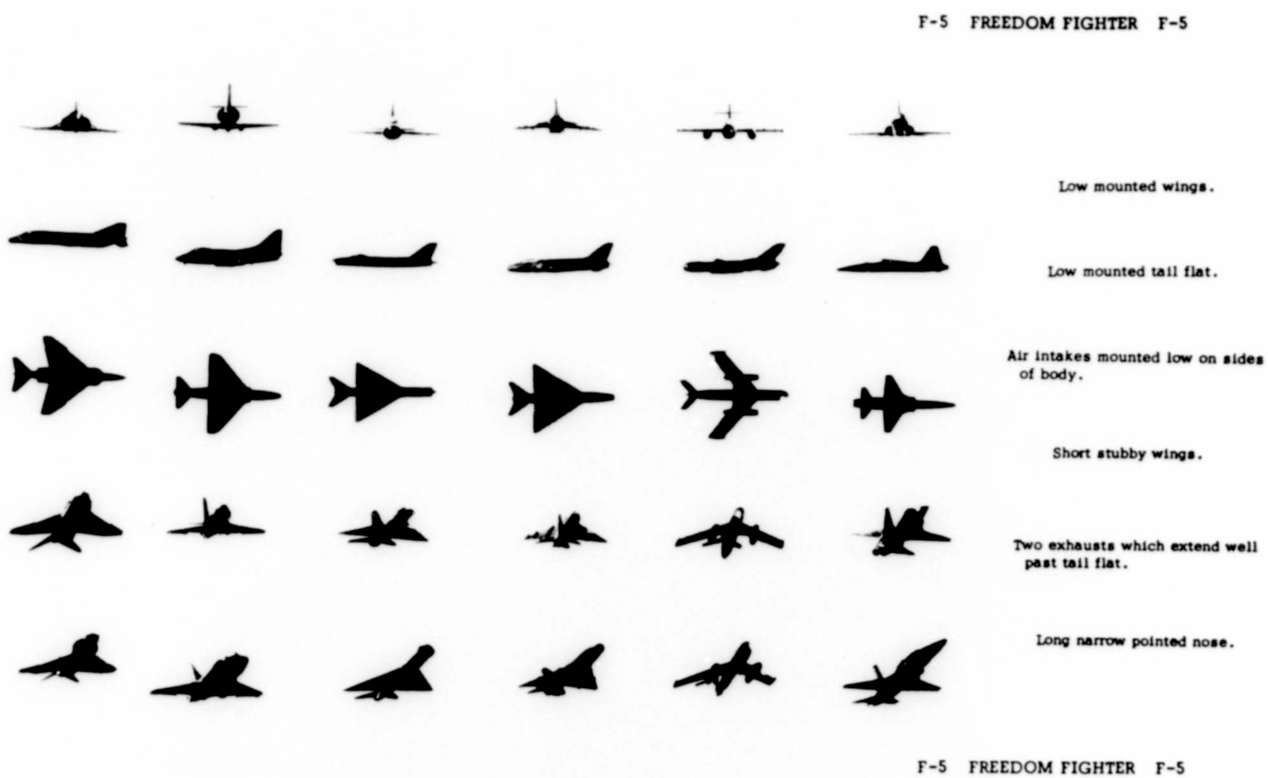


Figure A-2

HUMAN RESOURCES RESEARCH ORGANIZATION

300 North Washington Street • Alexandria, Virginia 22314

President	Dr. Meredith P. Crawford
Executive Vice President	Dr. William A. McClelland
Director for Business Affairs and Treasurer	Mr. Charles W. Smith
Director for Operations	Mr. Arnold A. Heyl
Director for Program Development	Dr. Robert G. Smith, Jr.
Director for Research Design and Reporting	Dr. Eugene A. Cogan

RESEARCH DIVISIONS

HumRRO Division No. 1 (System Operations) 300 North Washington Street Alexandria, Virginia 22314	Dr. J. Daniel Lyons Director
HumRRO Division No. 2 Fort Knox, Kentucky 40121	Dr. Donald F. Haggard Director
HumRRO Division No. 3 Post Office Box 5787 Presidio of Monterey, California 93940	Dr. Howard H. McFann Director
HumRRO Division No. 4 Post Office Box 2086 Fort Benning, Georgia 31905	Dr. T.O. Jacobs Director
HumRRO Division No. 5 Post Office Box 6057 Fort Bliss, Texas 79916	Dr. Albert L. Kubala Director
HumRRO Division No. 6 (Aviation) Post Office Box 428 Fort Rucker, Alabama 36360	Dr. Wallace W. Prophet Director
HumRRO Division No. 7 (Social Science) 300 North Washington Street Alexandria, Virginia 22314	Dr. Arthur J. Hoehn Director